American Journal of Otolaryngology–Head and Neck Medicine and Surgery xxx (xxxx) xxx

Contents lists available at ScienceDirect

American Journal of Otolaryngology–Head and Neck Medicine and Surgery

journal homepage: www.elsevier.com/locate/amjoto

How we ask matters: Ensuring accurate responses from AI in patient counseling

ARTICLE INFO

Keywords:
ChatGPT
Artificial intelligence
AI
Prompt engineering
Maxillofacial surgery
Otorhinolaryngology

Dear Editor,

We read with great interest the article by Mnajjed and Patel assessing the utility of ChatGPT-generated educational material for head and neck surgery counseling [1]. The study provided an insightful evaluation of ChatGPT's capabilities, particularly in comparison with traditional online educational resources. However, there are additional considerations that warrant discussion to fully appreciate the potential and limitations of this emerging technology in patient education.

First, while the study highlighted ChatGPT's comparable performance to existing online materials in providing basic perioperative information, it raises significant concerns about the bot's accuracy in identifying and advising on surgical complications. This limitation may be influenced by two factors. The first is the specific version of ChatGPT used, which the authors did not specify. The latest versions, such as GPT-4 and GPT-4o, have significantly improved their ability to provide more accurate and evidence-based responses [2–5]. Second, the accuracy of ChatGPT's responses can be enhanced by providing prompts that better contextualize the questions, allowing the AI to understand the role it needs to fulfill and the patient's specific needs [6,7]. While a physician naturally understands these aspects when counseling patients, the AI can only grasp them if explicitly instructed.

Moreover, the readability of ChatGPT's output, consistently rated at a college level, presents a barrier to accessibility. However, this aspect can also be improved by specifying in the prompt the type of language to use or the reading level of the intended recipient of the information [6–8]. By tailoring the language appropriately, the AI can generate content that is more accessible to patients with varying levels of health literacy.

Additionally, the article appropriately notes that ChatGPT's ability to interact with users when prompted is a unique feature that distinguishes it from traditional educational materials. However, it is also clear that none of the existing tools used to assess the quality of patient education materials, such as PEMAT or SAM, are fully applicable to chatbots like ChatGPT. These tools were designed to evaluate static content, whereas interactive AI systems require a different set of criteria to assess their effectiveness and safety accurately. It is therefore imperative that the healthcare community develops and validates new evaluation tools specifically tailored to the dynamic nature of AI-driven patient education platforms [9–11]. Such tools should assess not only the accuracy and readability of the information but also the quality of interaction, user engagement, and the system's ability to recognize and appropriately respond to complex medical scenarios.

In conclusion, while ChatGPT represents a significant advancement in AI-driven patient education, healthcare providers must remain vigilant about its limitations. AI should be used to augment, not replace, the nuanced care and judgment provided by experienced healthcare professionals.

CRediT authorship contribution statement

Luigi Angelo Vaira: Conceptualization, Writing – original draft, Writing – review & editing. **Giacomo De Riu:** Conceptualization, Writing – review & editing. **Carlos Miguel Chiesa-Estomba:** Conceptualization, Writing – review & editing. **Antonino Maniaci:** Conceptualization, Writing – review & editing. **Miguel Mayo-Yáñez:** Conceptualization, Writing – review & editing. **Alberto Maria Saibene:** Conceptualization, Writing – review & editing. **Jerome R. Lechien:** Conceptualization, Writing – review & editing.

Informed consent

N/A.

Ethical approval

N/A.

Funding

This work has been developed within the framework of the project e.INS-Ecosystem of Innovation for Next Generation Sardinia (cod. ECS 00000038) funded by Italian Ministry for Research and Education (MUR) under the National Recovery and Resilience Plan (NRRP) - MISSION 4 COMPONENT 2, "From research to business" INVESTMENT

https://doi.org/10.1016/j.amjoto.2024.104557
Received 24 August 2024
Available online 10 December 2024

1.5, "Creation and strengthening of Ecosystem of innovation" and construction of "Territorial R&D Leaders" CUP J83C21000320007.

Declaration of competing interest

None declared.

Acknowledgment

None declared.

References

[1] Mnajjed L, Patel RJ. Assessment of ChatGPT generated educational material for head and neck surgery counseling. Am J Otolaryngol 2024;45:104410.

[2] Lechien JR, Naunheim MR, Maniaci A, Radulesco T, Saibene AM, Chiesa-Estomba CM, et al. Performance and consistency of ChatGPT-4 versus otolaryngologists: a clinical case series. Otolaryngol Head Neck Surg 2024;170:1519–26.

[3] Lorenzi A, Pugliese G, Maniaci A, Lechien JR, Allevi F, Boscolo-Rizzo P, et al. Reliability of large language models for advanced head and neck malignancies management: a comparison between ChatGPT 4 and Gemini Advanced. Eur Arch Otorhinolaryngol May 25 2024. https://doi.org/10.1007/s00405-024-08746-2 [Epub ahead of print].

[4] Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, et al. Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: a multicenter collaborative analysis. Otolaryngol Head Neck Surg 2024;170:1492–503.

[5] Lechien JR, Briganti G, Vaira LA. Accuracy of ChatGPT-3.5 and -4 in providing scientific references in otolaryngology-head and neck surgery. Eur Arch Otorhinolaryngol 2024;281:2159–65.

[6] Campbell DJ, Estephan LE, Sina EM, Mastrolonardo EV, Alapati R, Amin DR, et al. Evaluating ChatGPT responses on thyroid nodules for patient education. Thyroid 2024;34:371–7.

[7] Vaira LA, Lechien JR, Abbate V, Gabriele G, Frosolini A, De Vito A, et al. Enhancing AI chatobot responses in healthcare: the SMART prompt structure in head and neck surgery. Res Sq 2024. https://doi.org/10.21203/rs.3.rs-4953716/v1.

[8] Momenaei B, Wakabayashi T, Shahlaee A, Durrani AF, Pandit SA, Wang K, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases. Ophthalmol Retina 2023;7:862–8.

[9] Lechien JR, Maniaci A, Gengler I, Hans S, Chiesa-Estomba CM, Vaira LA. Validity and reliability of an instrument evaluating the performance of intelligent chatbot: the Artificial Intelligence Performance Instrument (AIPI). Eur Arch Otorhinolaryngol 2024;281:2063–79.

[10] Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, et al. Validation of the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool: a new tool to assess the quality of health information provided by AI platforms. Eur Arch Otorhinolaryngol 2024;281:6123–31.

[11] Lavalle S, Lechien JR, Chiesa-Estomba C, Parisi FM, Maniaci A. Evaluating AI in patient education: the need for a validated performance assessment tool. Am J Otolaryngol 2024;45:104442.

Luigi Angelo Vaira[a,*], Giacomo De Riu[a], Carlos Miguel Chiesa-Estomba[b], Antonino Maniaci[c], Miguel Mayo-Yáñez[d], Alberto Maria Saibene[e], Jerome R. Lechien[f,g]

[a] *Maxillofacial Surgery Operative Unit, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Italy*

[b] *Department of Otorhinolaryngology-Head & Neck Surgery, Hospital Universitario Donostia, San Sebastian, Spain*

[c] *Department of Medicine and Surgery, University of Enna Kore, Enna, Italy*

[d] *Otorhinolaryngology, Head and Neck Surgery Department, Complexo Hospitalario Universitario A Coruña (CHUAC), A Coruña, Galicia, Spain*

[e] *Otolaryngology Unit, Santi Paolo e Carlo Hospital, Department of Health Sciences, University of Milan, Milan, Italy*

[f] *Department of Surgery, Mons School of Medicine, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium*

[g] *Department of Otolaryngology-Head Neck Surgery, Elsan Hospital, Paris, France*

[*] Corresponding author at: Maxillofacial Surgery Unit, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Viale San Pietro 10, 07100 Sassari, Italy.
E-mail address: lavaira@uniss.it (L.A. Vaira).